

# **PDS4 Product Search and Query Models**

J.S. Hughes, J.H. Padams, G.A.  
Hollins, and S.H. Hardman

Jet Propulsion Laboratory  
California Institute of Technology

4th Planetary Data Workshop

June 18–20, 2019

Flagstaff, Arizona

Planetary Data Archives

Support Services, and Tools I

Wednesday, June 19, 2019

9:45 a.m. HCCC Humphreys



**Jet Propulsion Laboratory**  
California Institute of Technology



# Search Goals

- **Improved integration across nodes and across agencies**
  - *Pass search parameters to the node search engines (e.g., two/n-tiered search).*
  - *International data sets integrated with PDS search results.*
- **Better navigation and support for mission and other “virtual” views**
  - *resources that can be browsed / linked-to easily*
  - *PDS mission support pages registered in the PDS search infrastructure.*
  - *Facet-based navigation to drive specific views.*



# Development Approach

- **Build the software system based on generic common software and common protocols for accessing that software.**
  - *PDS Registry along its access APIs is the main component.*
  - *PDS Registry based on Apache Solr provides support for high performance facet-based search.*
- **Utilize the PDS4 data model for data object definitions and to configure the software where appropriate.**
  - *The model defines the key context objects (i.e., Investigation, Instrument, Collection, Observational Data, etc.).*



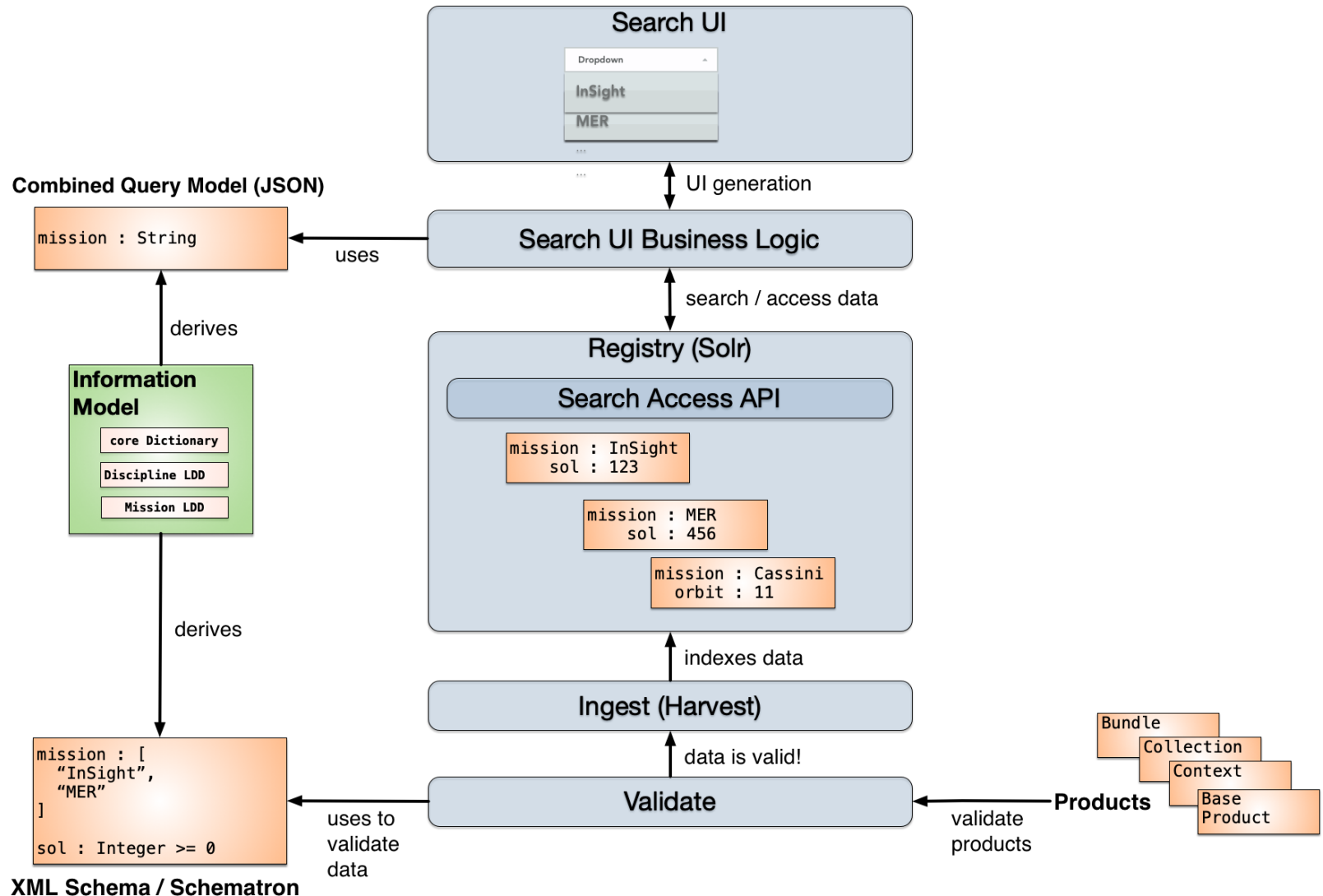
# Registry

- **This service is a deployable component that accepts queries for data and returns a set of matching results.**
- **Provides the public interface (REST-based API over HTTP) to the metadata contained in the federated instances of the Registry Service.**
- **Generation of search indices from registry metadata supports multiple query formats and is tailor-able for customized search interfaces.**





# Search Architecture





# Search Protocol

- **Common PDS-based parameters**
  - *Includes identifier, instrument, investigation, etc.*
  - *Includes mapping from PDS3 to PDS4.*
- **Supports simple and advanced syntaxes**
  - *Simple syntax closely follows HTTP.*
  - *Advanced syntax closely follows Apache Lucene/Solr.*
- **Supports multiple result set formats**
  - *Default format available in XML and JSON.*
  - *Initial support for VOTable.*
- **All the above detailed in the PDS Search Protocol [1] and PDAP Search Protocol [2] documents.**

[1] [https://pds.nasa.gov/services/pds4\\_pds\\_search\\_protocol.pdf](https://pds.nasa.gov/services/pds4_pds_search_protocol.pdf)

[2] [https://pds.nasa.gov/services/pds4\\_pdap\\_search\\_protocol.pdf](https://pds.nasa.gov/services/pds4_pdap_search_protocol.pdf)



# Service Configuration

- **Much of the functionality provided by the Registry and Search is achieved through manual configuration.**
  - *Defining search fields, facets, etc. in Apache Solr.*
  - *Mapping XPath to **human-readable** searchable fields*
- **Wouldn't it be great if this could be automatically configured directly from the PDS Information Model?**
  - *Query models appear to be the answer.*



# What is the purpose of a Query Model?

- **A query model identifies attributes that are useful for search.**
  - *PDS4 product labels contain information that is useful for locating data products in the archive.*
- **A query model is used to configure system functions**
  - *Harvester extracts the value of the identified attributes*
  - *Search API to allow searching these values*



# Product Label Template

Identification_Area Logical_Identifier Version_Id	
Observation_Area Time_Coordinates Primary_Result_Summary Investigation_Area Observing_System Target_Identification	<i>Discipline_Area</i> <i>Mission Area</i>
Reference_List Internal_Reference External_Reference	
File_Area_Observational File Header Array_2D_Image	



# Some Useful Search Attributes

## Target: Product Collections (Mission Science Data Collections)

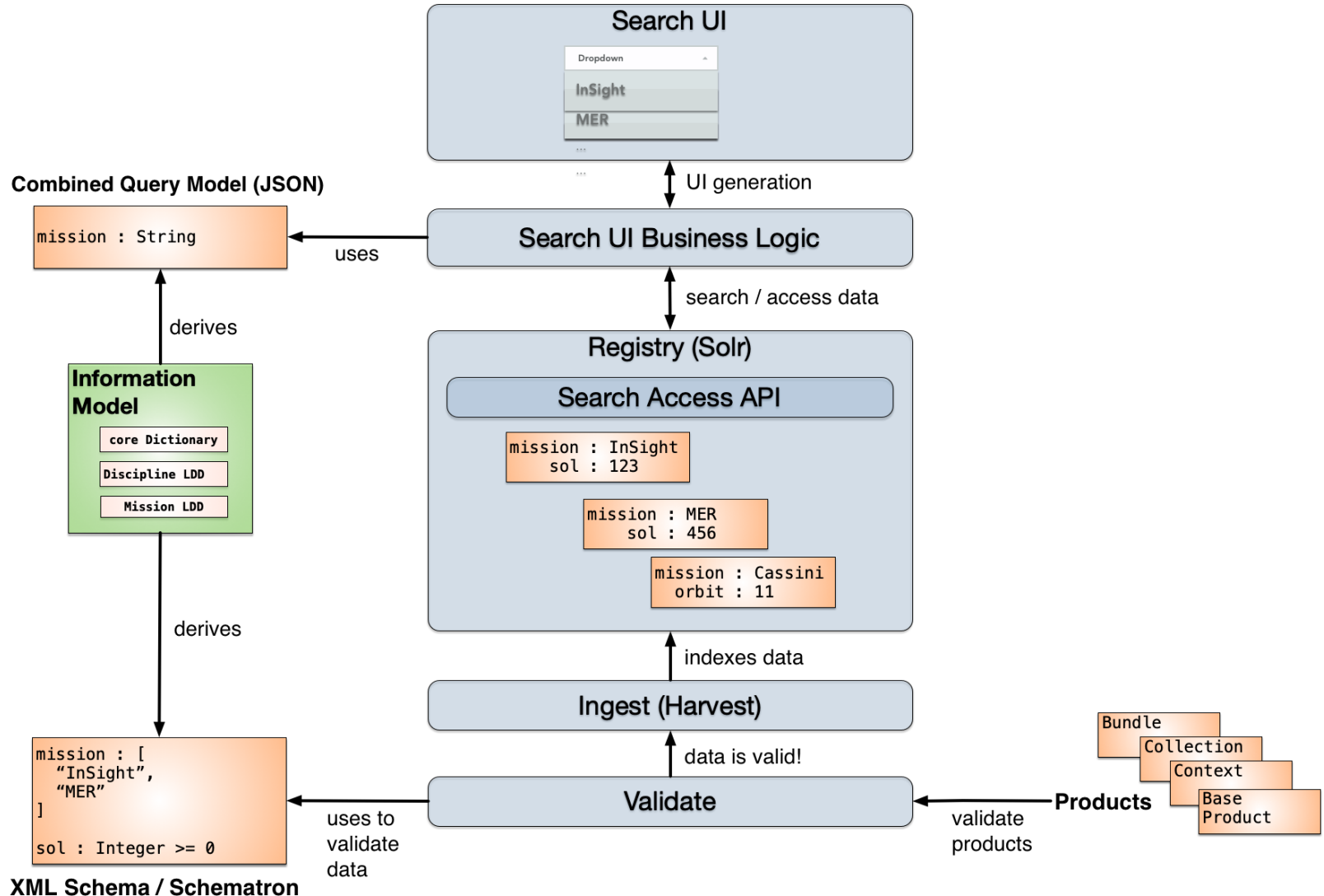
- **collection\_type=Data**
- **type=Mission**
- **product\_class=Product\_Collection**
- **purpose=Science**

Search Parameters for Mission Science Data Collections		
Candidate attribute for search	Faceted Search	Text- or Field-Search
Product_Collection.Collection.collection_type=Data	Yes	Yes
Investigation_Area.type=Mission	Yes	Yes
Identification_Area.product_class=Product_Collection	Yes	Yes
Primary_Result_Summary.purpose=Science	Yes	Yes
Time_Coordinates.start_date_time		Yes
Time_Coordinates.stop_date_time		Yes
Target_Identification.name		Yes
Target_Identification.type	Yes	Yes
Investigation_Area.name		Yes
Observing_System.name		Yes
Observing_System_Component.name		Yes
Observing_System_Component.type (type=Instrument)	Yes	Yes
Primary_Result_Summary.processing_level	Yes	Yes
Science_Facets.domain	Yes	Yes
Science_Facets.wavelength_range	Yes	Yes
Science_Facets.discipline_name	Yes	Yes
Collection.description		Yes
Product_Context.Instrument.type	Yes	Yes
Identification_Area.logical_identifier		Yes
Identification_Area.version_id		Yes
Science_Facets.facet1	Yes	Yes
Science_Facets.facet2	Yes	Yes
Primary_Result_Summary.description		Yes
Target_Identification.description		Yes
Identification_Area.title		Yes
system.archive_status	Yes	Yes
Citation_Information.keyword		Yes
Citation_Information.description		Yes





# Search Architecture





# Product Label Template

Identification_Area	
Logical_Identifier	
Version_Id	
Observation_Area	
Time_Coordinates	
Primary_Result_Summary	<i>Discipline_Area</i>
Investigation_Area	<i>Mission Area</i>
Observing_System	
Target_Identification	
Reference_List	
Internal_Reference	
External_Reference	
File_Area_Observational	
File	
Header	
Array_2D_Image	



# How are Query Models populated?

<Ingest\_LDD

<name>Cartography</name>

<DD\_Attribute>

<name>map\_projection\_name</name>

<version\_id>1.0</version\_id>

<local\_identifier>map\_projection\_name</local\_identifier>

<nillable\_flag>>false</nillable\_flag>

<query\_model\_name>Cart QM 1</query\_model\_name>

<query\_model\_name>Cart QM 2</query\_model\_name>

<submitter\_name>Elizabeth D. Rye</submitter\_name>

<definition>The map\_projection\_name attribute provides ... </definition>

<DD\_Value\_Domain>

**Thank You**



# Data Flow

